ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 6, June 2017

Entropy Optimized Feature-Based Bag-Of-Words Representation for Information Retrieval

Swaroop A.Kale[1], Prof. H.A.Hingoliwala[2]

M. E Student, Department of Computer Engineering, Jaywantrao Sawant College of Engineering, Pune, India.[1]

Head of Dept, Department of Computer Engineering, Jaywantrao Sawant College of Engineering, Pune, India[2]

ABSTRACT: Information retrieval (IR) is the activity of obtaining information resources relevant to an information need from a collection of information resources. Searches can be based on full-text or other content-based indexing. Feature quantization is a crucial component for efficient large scale image retrieval and object recognition. By quantizing local features into visual words, one hope that features that match each other obtain the same word ID. Then, similarities between images can be measured with respect to the corresponding histograms of visual words. Given the appearance variations of local features, traditional quantization methods do not take into account the distribution of matched features. In existing works, gain of the optimization is less obvious due to the difference of the distribution of training and test data. To tackle this existing gain problem, we proposed, Supervised Feature Quantization with Entropy Optimized Feature-based Bag-of-Words Representation for Information Retrieval. Feature quantization is a crucial component for efficient large scale image retrieval and object recognition. By quantizing local features into visual words, one hope that features that match each other obtain the same word ID. Then, similarities between images can be measured with respect to the corresponding histograms of visual words. Given the appearance variations of local features, traditional quantization methods do not take into account the distribution of matched features. In this system, we investigate how to encode additional prior information on the feature distribution via entropy optimization by leveraging ground truth correspondence data. We propose a computationally efficient optimization scheme for large scale vocabulary training.

KEYWORDS: Information retrieval (IR), Information Retrieval, Bag-of-Features (BoF) or Bag-of-Visual Words (BoVW).

I. INTRODUCTION

Information retrieval (IR) is the task of retrieving objects, e.g., images, from a database given the user's information need. Early research focused mostly on text retrieval, but then quickly expanded to other areas, such as image retrieval, and video retrieval, since the availability of digital technology led to a great increase of multimedia data.

A similar growth in the availability of time series data, i.e., data that are composed of a sequence of measurements, such as medical monitoring data, also raised the interest in the retrieval of time series. We can more formally define IR as follows. Given a collection of objects D = {d1, d2,..., dn} and a query object q, rank the objects in D according to their relevance to q and select a subset of the most relevant objects. As we already mentioned, the collection can contain any type of objects, such as images, videos, time-series or text documents. However, we focus mainly on image retrieval since it is the most studied and challenging aspect of multimedia information retrieval. This is without loss of generality as the proposed method can be applied to several other types of data, such as video, audio and time-series, with minor modifications.

The proposed method cannot be directly used for text retrieval, however it can utilize word-embedding techniques, to create optimized representations for text documents. One of the most important challenges in image retrieval is the so-called semantic gap, between the low-level representation of an image and its higher level concepts. In other words, the semantic gap describes the phenomenon in which images with similar semantic content have very different low-

DOI: 10.15680/IJIRCCE.2017. 0506025

level representations (e.g., color information) and vice versa. The high dimensionality of images reduces the retrieval performance even more, both in terms of query time and precision. Several approaches have been proposed to extract meaningful low-dimensional features from images.

The purpose of feature extraction is to lower the dimensionality of the images, which reduces the storage requirements and the retrieval time, and to bridge the semantic gap, which increases the retrieval precision. Perhaps the most widely used and successful method for this task is the feature-based Bag-of-Words model, also known as Bag-of-Features (BoF) or Bag-of-Visual Words (BoVW).

The feature-based BoW approaches, should not be confused with the standard textual BoW approaches, that are discussed later on in this Section. In the rest of the manuscript we abbreviate the feature-based Bag-ofWords models as BoW. The BoW model treats each image as a document that contains a number of different “visual” words. Then an image is represented as a histogram over a set of representative words, known as dictionary or codebook. These histograms describe the corresponding images and they can be used for the subsequent retrieval tasks. The BoW pipeline can be summarized as follows.

1) feature extraction, in which multiple features, such as SIFT descriptors, are extracted from each image. That way, the feature space is formed where each image is represented as a set of features.

2) dictionary learning, in which the extracted features are used to learn a dictionary of representative features (also called words or codewords),

3) feature quantization and encoding, in which each feature is represented using a codeword from the learned dictionary and a histogram is extracted for each image. That way, the histogram space is formed where each image is represented by a constant dimensionality histogram vector.

Unlike textual words, visual words are not predefined and the quality of the extracted representation critically relies on how these words are chosen. Early feature-based BoW approaches used unsupervised clustering algorithms, such as k-means, to cluster the set of features and learn a dictionary. These unsupervised approaches achieved promising results and produced codebooks that were generic enough to be used for any task. However, learning a discriminative dictionary tailored to a specific problem is expected to perform significantly better. Indeed, supervised dictionary learning, is shown to achieve superior performance in terms of classification accuracy. These methods produce discriminative dictionaries that are useful for the given classification problem.

Even though a highly discriminative representation is desired for classification tasks, it is not always optimal for retrieval since it might severely distort the similarity between images in order to gain discrimination ability. This can be better understood by an example. Suppose that we want to learn a dictionary that distinguishes apples from oranges using only two code words and a perfect discriminative dictionary learning algorithm exists. After the training process, each apple is represented by a vector (x, y) is belong into N(0,0.1) × N(1,0.1) and each orange by a vector (x, y) is belong into N(1,0.1) × N(0,0.1), where N(μ, σ^2) is a normal distribution with mean μ and variance σ^2.

This codebook would be excellent for classification and retrieval of oranges and apples, as it manages to linearly separate the two classes with a large margin. Now, what would happen if we use this representation to retrieve an another fruit, such as pears or bananas? Does this dictionary gained its discrimination ability at the expense of its representation ability and would actually perform worse (than an unsupervised dictionary) on this task? Our experiments confirm this hypothesis, since we found that a highly discriminative representation excels at its learned domain, but its discriminative ability outside this domain is severely limited. The interested reader, where the problems of transfer learning and domain adaptation are discussed. We also expect relevance feedback techniques, i.e., methods that are used to better identify the user’s information need, not to work correctly outside the training domain, since the representation ability is already lost.

Ideally, the aforementioned problem would be solved if we could optimize the learned representation for every possible information need. Since this is rather infeasible, we can learn a representation using a large-scale database with annotated images, such as ImageNet, that covers a broad range of information needs. However, it is not always possible to acquire a large set of annotated training data, mostly because of the high cost of annotation. Therefore, a good representation for retrieval should

a) improve the retrieval metrics inside its training domain and

b) be able to “transfer” the learned knowledge to other similar domains.

The latter is especially important as it ensures that we can learn using a small and representative training set. To this end, we propose a supervised dictionary learning method that produces retrieval-oriented codebooks by adhering to the cluster hypothesis. Cluster hypothesis states that points in the same cluster are likely to fulfill the same information need. We select a set of centroids in the histogram space and we learn a codebook that minimizes the entropy of each cluster. Each centroid can be considered as a representative query and the optimization aims to maximize the relevant information around it.

The entropy objective acts as a mild discrimination criterion, which try to make the clusters as pure as possible. To understand why this criterion differs from other more discriminative criteria, such as the Fisher's ratio, or max-margin objectives, note that we do not push clusters or points away from each other. Instead, the clusters are fitted to the existing data distribution and we only try to move irrelevant points to the nearest relevant cluster. This allows us to optimize the codebook without over-fitting the representation over the training domain.

We validate our claims by demonstrating the ability of the proposed method to successfully retrieve images that belong to classes that are not seen during the training process. We should mention that our method is not limited to image retrieval. It is general enough to be applied to any task that involves BoW representations, such as video, audio, and time series retrieval. The term Bag-of-Words is also used to refer to the natural language processing methods that handle a text document as collection of its words.

One of the most widely used such methods uses a term frequency (tf) histogram, which counts the appearances of each word, to represent a document as a vector. In these approaches the words of the dictionary are predefined and cannot be altered. Dictionary learning for this representation aims mainly to prune the dictionary by selecting the most useful features, instead of altering the words and extracting a new representation. Although the proposed method cannot be applied when the textual BoW representation is used, with the advent of the word-embedding models, e.g., is possible to map each word to a "meaningful" low-dimensional continuous vector. Then, the proposed method can be used by extracting a feature vector from each word of a text document.

A similar approach is used to apply the proposed method to a dataset that contains movie reviews in the form of free text. The main contribution of this paper is the proposal of a novel retrieval-oriented dictionary learning method which can a) improve the retrieval precision over the state-ofthe-art methods, b) reduce the storage requirements and query time by using smaller dictionaries, and c) transfer the learned knowledge to previously unseen classes without retraining.

II. RELATED WORK

S. Lazebnik and M. Raginsky [2], " A Supervised learning of quantizer codebooks by information loss minimization ", proposed In most of bag of words image classification methods bag of features form the basic attribute for comparison. To achieve bag of features in compressed form quantization is used that compresses the high level features of images in to small codeword's.Authors have considered the problem of quantizing continuous feature spaces while preserving structure that is necessary for predicting a given target attribute. The basic idea behind our method is that the compressed representation of the data should be a sufficient statistic for the attribute, i.e., it should preserve all information about that attribute. Here the optimization aims to increase the mutual information between each codeword and the feature labels so from a code word we can effectively derive correct features which will help in classification of images. Authors have only addressed the relatively simple scenario where the compression is accomplished by nearest-neighbor quantization and the target task is to predict a discrete label in future they would consider scenario for wider class of problems of task-specific compression.

Y. Kuang, M. Byrod, and K. Astrom [3] A Supervised feature quantization with entropy optimization

has proposed that for image categorization, the aim of supervised feature quantization is to incorporate semantic categorical information into the training vocabulary in such a way that the histogram representation of images encodes the patterns of each category more accurately. For local feature matching, such a similarity measure is generally a proper criterion. However, due to lighting conditions, perspective transformation, etc. local features can be very different from each other. In this case, unsupervised feature quantization based solely on similarity might fail to capture the intra-class variation of local features. Supervised feature quantization on the other hand utilizes correspondence labels and improves matching performance with respect to such intra-class variation. Authors study a supervised feature quantization approach based on entropy optimization. Disadvantage of these methods is that each feature carries the label of the image in which it appears. This assumption does not always hold and it can negatively affect the quality of the learned codebook. For example, a sky patch may both appear in a sea scene and in a forest scene and it would consider both images in to same category it would be wrong to assume that.

X.-C. Lian, Z. Li, B.-L. Lu, and L. Zhang et al., [4] proposed “A Max-margin dictionary learning for multiclass image categorization” shows Visual dictionary learning and base (double) classifier preparing are two essential issues for the as of late most well known picture order structure, which depends on the sack of-visual-terms (BOV) models. Existing system normally manage the over two issues separately: dictionaries are initially created and classifiers are then learned based on them. In this paper, we propose a novel method named Max-Margin Dictionary Learning(MMDL) for image Classification which binds together the word reference learning process with classifier preparing. The structure lessens the multi-class issue to a gathering of one-versus one parallel issues. For every binary problem , classifier learning and dictionary generation are directed iteratively by minimizing a brought together target work which embraces the greatest edge criteria.

H. J´egou, M. Douze, C. Schmid, and P. P´erez [5], " A Aggregating local descriptors into a compact image representation", It proposes that in many of existing works of image retrieval based ob Bag of Features (BOF) model of image classification indexing was a mechanism which would improve the performance of image retrieval system. But disadvantage of the system was that indexing would take a large of storage as well that affects the performance when database is large. This motivated Authors to device a descriptor termed VLAD (vector of locally aggregated descriptors) , derived from both BOF and Fisher kernel , that aggregates SIFT descriptors and produces a compact representation thus saving a lot of space requirement hence improving the performance of system as it was cheaper to compute.

As per Andoni and P. Indyk et al., [6] has examined that the goal of this article is twofold. In the first part, we survey a family of nearest neighbor algorithms that are based on the concept of locality sensitive hashing. Many of these algorithm have already been successfully applied in a variety of practical scenarios. In the second part of this article, we describe a recently discovered hashing-based algorithm, for the case where the objects are points in the d-dimensional Euclidean space. As it turns out, the performance of this algorithm is provably near-optimal in the class of the locality-sensitive hashing algorithms.

III. PROPOSED SYSTEM

To tackle this existing gain problem, we proposed, Supervised Feature Quantization with Entropy Optimized Feature-based Bag-of-Words Representation for Information Retrieval. Feature quantization is a crucial component for efficient large scale image retrieval and object recognition. By quantizing local features into visual words, one hopes that features that match each other obtain the same word ID. Then, similarities between images can be measured with respect to the corresponding histograms of visual words. Given the appearance variations of local features, traditional quantization methods do not take into account the distribution of matched features. In this system, we investigate how to encode additional prior information on the feature distribution via entropy optimization by leveraging ground truth correspondence data. We propose a computationally efficient optimization scheme for large scale vocabulary training.

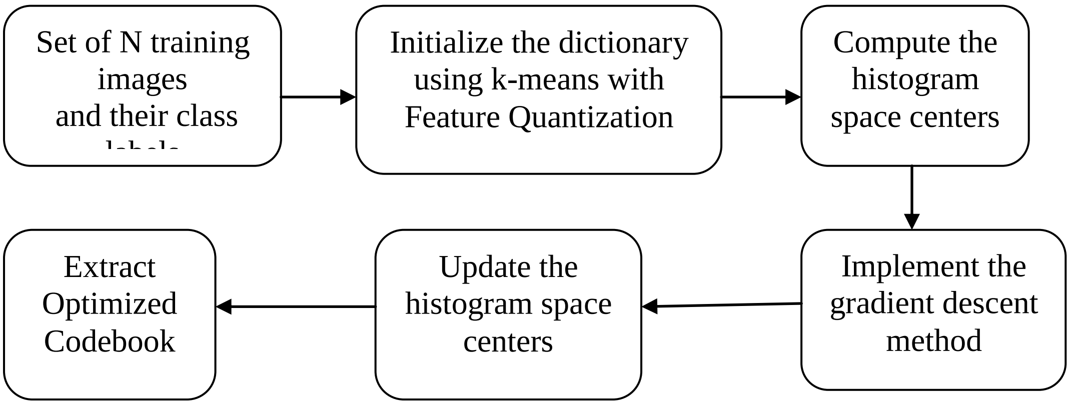
Set of N training images and their class labels
Initialize the dictionary using k-means with Feature Quantization
Compute the histogram space centers
Implement the gradient descent method
Update the histogram space centers
Extract Optimized Codebook

Fig. 1. System Architecture

Load Training Images:
In this module we load the set of N training images and their class labels. The goal of the EO-BoW technique is to learn a codebook that minimizes the entropy in the histogram space by using a training set of images, where the i-th image is annotated by a label li ∈ {1, ..., NC} and NC is the number of training classes.

Initialize the dictionary using k-means with Feature Quantization:
Intuitively, the entropy in the histogram space is minimized when the image histograms are gathered in pure clusters, i.e., each cluster contains images of the same class. Then we initialize the dictionary using k-means. The k-means algorithm takes as input the number of clusters to generates k, and a set of observation vectors to cluster. It returns a set of centroids, one for each of the k clusters. An observation vector is classified with the cluster number or centroid index of the centroid closest to it. Then we apply feature quantization. Since feature quantization is a natural application for k-means, information theory terminology is often used. The centroid index or cluster index is also referred to as a “code” and the table mapping codes to centroids and vice versa is often referred as a “code book”. The result of k-means, a set of centroids, can be used to quantize features. Quantization aims to find an encoding of features that reduces the expected distortion.

Extract Optimized Codebook:
In this module, we compute the histogram space centers by separately running k-means for each class. Then we implement the gradient descent method with backtracking. After each iteration we update the histogram space centers by running k-means again. Finally we extract optimized retrieval-oriented codebooks.

Design Considerations:

- Before requesting for Private Key owner must have generated Pseudonym.
- Pseudonym Generated is always of fixed length and has no connection with name / identity of user who generates it.
- Pseudonym is the identity of user for CSP and PKG both.
- Attributes used to decide the access policies.
- Attributes will be used to decide who can decrypt the Document.
- Decrypt the document only if access policy is satisfied with set of user attributes.

Description of the Proposed Algorithm:

Supervised Feature Quantization with EO-BOW Algorithm
Input: A set of N training images and their class labels
Output: The optimized codebook V
1) First, we initialize the dictionary using k-means. The k-means algorithm takes as input the number of clusters to generates k, and a set of observation vectors to cluster. It returns a set of centroids, one for each of the k clusters. An observation vector is classified with the cluster number or centroid index of the centroid closest to it.

2) Then we apply feature quantization. Since feature quantization is a natural application for k-means, information theory terminology is often used. The centroid index or cluster index is also referred to as a “code” and the table mapping codes to centroids and vice versa is often referred as a “code book”. The result of k-means, a set of centroids, can be used to quantize features. Quantization aims to find an encoding of features that reduces the expected distortion.
3) Then, we compute the histogram space centers by separately running k-means for each class.

4) Then we implement the gradient descent method with backtracking.

5 After each iteration we update the histogram space centers by running k-means again.

6)) Finally we extract an optimized retrieval-oriented codebooks.

Pseudo code :

```
Step 1: Determine the number of codevectors N ⍰
Step 2:  Select N codevectors at random to be the initial codebook
Step 3: Select N codevectors at random to be the initial codebook
Step 4: Compute the new set of codevectors (codebook)
Step 5: Repeat Steps 2 and 3 until the either the representative codevectors do not change
Step 6: End.
```

IV. Simulation Results

We proposed, Supervised Feature Quantization with Entropy Optimized Feature-based Bag-of-Words Representation for Information Retrieval. Feature quantization is a crucial component for efficient large scale image retrieval and object recognition. By quantizing local features into visual words, one hopes that features that match each other obtain the same word ID. Then, similarities between images can be measured with respect to the corresponding histograms of visual words. Given the appearance variations of local features, traditional quantization methods do not take into account the distribution of matched features. In this system, we investigate how to encode additional prior information on the feature distribution via entropy optimization by leveraging ground truth correspondence data. We propose a computationally efficient optimization scheme for large scale vocabulary training.

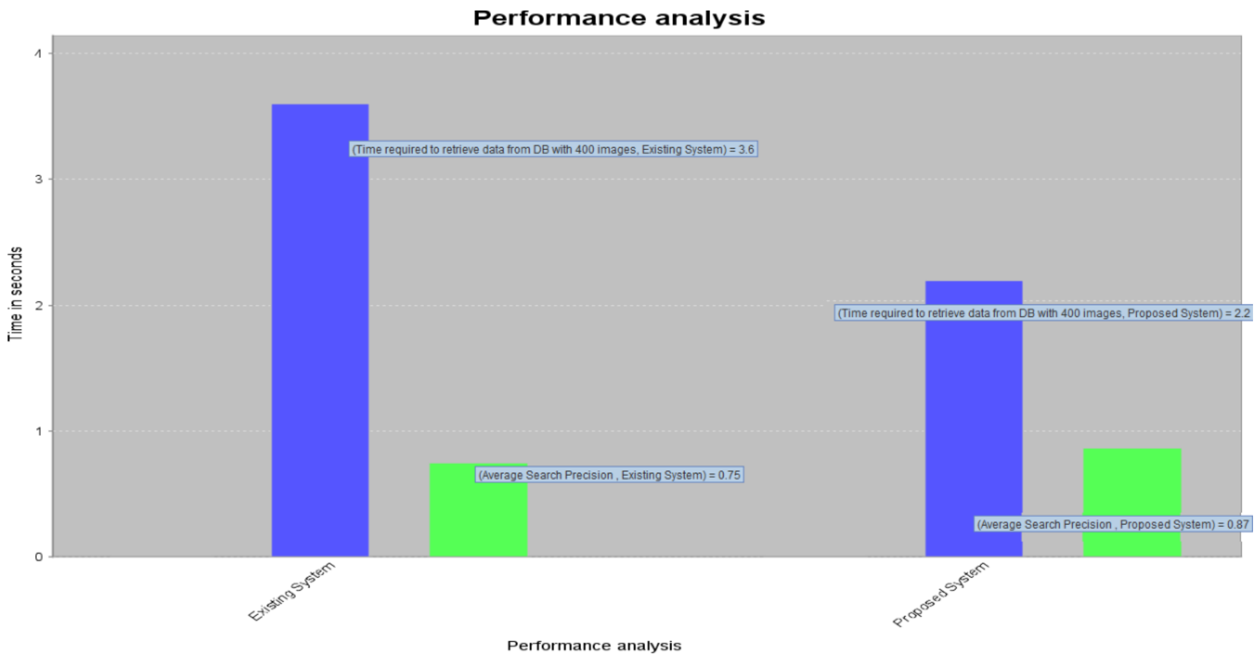

Fig. 1. Performance analysis

Figure 1 shows blue color bar is for Time required to retrieve data from DB with N images
where N = 400
Time is in seconds

green color for Average Search Precision between 0.1 to 1.0
0.7 stands for 70% images relevant retried

V. Conclusion and Future Work

The proposed In this paper we proposed a supervised dictionary learning method, the EO-BoW, which optimizes a retrieval-oriented objective function. We demonstrated the ability of the proposed method to improve the retrieval performance using two image datasets, a collection of time-series datasets, a text dataset and a video dataset. First, for a given dictionary size, it can improve the mAP over the baseline methods and other state-of-the-art representations. Second, these improvements allow us to use smaller representations which readily translates to lower storage requirements and faster retrieval. In exchange for these, our method requires a small set of annotated training data. Although the gained performance is correlated to the size and the quality of the training dataset, we showed that the proposed method does not lose its representation ability even when a small training dataset is used. Finally, we demonstrated that the EOBoW improves the retrieval performance using two different similarity metrics, the euclidean and the chi-square distance. Therefore, it can be combined with any approximate nearest neighbor technique that works with these similarity metrics to further increase the retrieval speed.

Currently proposed work is not always possible to acquire a large set of annotated training data, mostly because of the high cost of annotation. In future, we tackle this problem.

References

1. A. Andoni and P. Indyk, “Near-optimal hashing algorithms for approximate nearest neighbor in high dimensions,” IEEE Symposium on Foundations of Computer Science, Vol.11,pp. 459–468, 2006.
2. M. M. Baig, H. Gholamhosseini, and M. J. Connolly, “A comprehensive survey of wearable and wireless ECG monitoring systems for older adults,” Medical & Biological Engineering & Computing, vol. 51, pp. 485–495, 2013.
3. Y.-L. Boureau, F. Bach, Y. LeCun, and J. Ponce, “Learning mid-level features for recognition,” in IEEE Conference on Computer Vision and Pattern Recognition, Vol.11, pp. 2559–2566,2010.
4. F. K.-P. Chan, A. W.-C. Fu, and C. Yu, “Haar wavelets for efficient similarity search of time-series: with and without time warping,” IEEE Transactions on Knowledge and Data Engineering, vol. 15, pp. 686–705, 2003.
5. D. Chatzakou, N. Passalis, and A. Vakali, “Multispot: Spotting sentiments with semantic aware multilevel cascaded analysis,” in Big Data Analytics and Knowledge Discovery,Vol.41,, pp. 337–350,2015.
6. R. Datta, D. Joshi, J. Li, and J. Z. Wang, “Image retrieval: Ideas, influences, and trends of the new age,” ACM Computing Surveys, vol. 40, pp. 2559–2566, 2008.
7. D. Gorisse, M. Cord, and F. Precioso, “Locality-sensitive hashing for chi2 distance,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 34, pp. 402–409, 2012.
8. S. C. Hoi, M. R. Lyu, and R. Jin, “A unified log-based relevance feedback scheme for image retrieval,” IEEE Transactions on Knowledge and Data Engineering, vol. 18, pp. 509–524, 2006.
9. A. Iosifidis, A. Tefas, and I. Pitas, “Multidimensional sequence classification based on fuzzy distances and discriminant analysis,” IEEE Transactions on Knowledge and Data Engineering, vol. 25, pp. 2564–2575, 2013.

BIOGRAPHY

Miss Swaroop Arun Kale is currently pursuing M.E. (Computer Engineering), Jaywantrao Sawant College of engineering, Pune, Maharashtra, India – 411028. She received her B.E. (Computer and science) Degree from M.S.Bidve College of engineering, Latur, Maharashtra, India-413512. Her area of interest is Information Retrieval and Data Mining.